

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Sullivan, Francis
Kriz, Ron
Kumar, Ravindra

(ii) TITLE OF INVENTION: Human GDP-Mannose Hydratase

(iii) NUMBER OF SEQUENCES: 5

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Genetics Institute, Inc.
(B) STREET: 87 CambridgePark Drive
(C) CITY: Cambridge
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02140

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Brown, Scott A.
(B) REGISTRATION NUMBER: 32,724
(C) REFERENCE/DOCKET NUMBER: GI5285

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (617) 498-8224
(B) TELEFAX: (617) 876-5851

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1521 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCCGCGTTCC TGCCGGCACC GCGCCTGCCC TCTGCCGCGC TCCGCCCTGC CGCCGACCGC	60
ACGCCCGCCG CGGGACATGG CACACGCACC GGCACGCTGC CCCAGCGCCC GGGGCTCCGG	120
GGACGGCGAG ATGGGCAAGC CCAGGAACGT GCGCTCATC ACCGGTATCA CAGGCCAGGA	180

TGGTTCCTAC CTGGCTGAGT TCCTGCTGGA GAAAGGCTAT GAGGTCCATG GAATTGTACG	240
GCGGTCCAGT TCATTTAATA CGGGTCGAAT TGAGCATCTG TATAAGAATC CCCAGGCTCA	300
CATTGAAGGA AACATGAAGT TGCACTATGG CGATCTCACT GACAGTACCT GCCTTGTGAA	360
GATCATTAAAT GAAGTAAAGC CCACAGAGAT CTACAACCTT GGAGCCCAGA GCCACGTCAA	420
AATTCCTTT GACCTCGCTG AGTACACTGC GGACGTTGAC GGAGTTGGCA CTCTACGACT	480
TCTAGATGCA GTTAAGACTT GTGGCCTTAT CAACTCTGTG AAGTTCTACC AAGCCTCAAC	540
AAGTGAACCTT TATGGGAAAG TGCAGGAAAT ACCCCAGAAG GAGACCACCC CTTTCTATCC	600
CCGGTCACCC TATGGGGCAG CAAACTCTA TGCCTATTGG ATTGTGGTGA ACTTCCGTGA	660
GGCGTATAAT CTCTTTGCAG TGAACGGCAT TCTCTTCAAT CATGAGAGTC CCAGAAGAGG	720
AGCTAATTC GTTACTCGAA AAATTAGCCG GTCAGTAGCT AAGATTTACC TTGGACAAC	780
GGAAATGTTT AGTTTGGGAA ATCTGGATGC CAAACGAGAT TGGGGCCATG CCAAGGACTA	840
TGTGGAGGCT ATGTGGTTGA TGTTCAGAA TGATGAGCCG GAGGACTTCG TTATAGCTAC	900
TGGGGAGGTC CATAGTGTC GGGAAATTTGT CGAGAAATCA TTCTGCACA TTGAAAAAC	960
CATTGTGTGG GAAGGAAAGA ATGAAATGA AGTGGGCAGA TGTAAGAGA CCGGCAAAGT	1020
TCACGTGACT GTGGATCTCA AGTACTACCG GCCAACTGAA GTGGACTTTC TGCAGGGCGA	1080
CTGCACCAA GCGAAACAGA AGCTGAACTG GAAGCCCCGG GTCGCTTTCG ATGAGCTGGT	1140
GAGGGAGATG GTGCACGCCG ACGTGGAGCT CATGAGGACA AACCCCAATG CCTGAGCAGC	1200
GCCTCGGAGC CCGGCCGCC CTCCGGCTAC AATCCCCGCA GAGTCTCCGG TGCAGACGCC	1260
CTGCGGGGAT GGGGAGCGGC GTGCCAATCT GCGGTCCCC TCGGGCCCT GCTGCCGCTG	1320
CGCTGTCCCG GCCGCAAGAG CGGGGCCGCC CCGCCGAGGT TTGTAGCAGC CGGGATGTGA	1380
CCCTCCAGGG TTTGGGTCGC TTTGCGTTTG TCGAAGCCTC CTCTGAATGG CTTTGTGAAA	1440
TCAAGATGTT TTAATCACAT TCACTTTACT TGAAATTATG TTGTTACACA ACAAATTGTG	1500
GGGCCTTCAA ATGTTTTTTC C	1521

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 354 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Gly Lys Pro Arg Asn Val Ala Leu Ile Thr Gly Ile Thr Gly Gln

1	5	10	15
Asp Gly Ser Tyr Leu Ala Glu Phe Leu Leu Glu Lys Gly Tyr Glu Val	20	25	30
His Gly Ile Val Arg Arg Ser Ser Ser Phe Asn Thr Gly Arg Ile Glu	35	40	45
His Leu Tyr Lys Asn Pro Gln Ala His Ile Glu Gly Asn Met Lys Leu	50	55	60
His Tyr Gly Asp Leu Thr Asp Ser Thr Cys Leu Val Lys Ile Ile Asn	65	70	75
Glu Val Lys Pro Thr Glu Ile Tyr Asn Leu Gly Ala Gln Ser His Val	85	90	95
Lys Ile Ser Phe Asp Leu Ala Glu Tyr Thr Ala Asp Val Asp Gly Val	100	105	110
Gly Thr Leu Arg Leu Leu Asp Ala Val Lys Thr Cys Gly Leu Ile Asn	115	120	125
Ser Val Lys Phe Tyr Gln Ala Ser Thr Ser Glu Leu Tyr Gly Lys Val	130	135	140
Gln Glu Ile Pro Gln Lys Glu Thr Thr Pro Phe Tyr Pro Arg Ser Pro	145	150	155
Tyr Gly Ala Ala Lys Leu Tyr Ala Tyr Trp Ile Val Val Asn Phe Arg	165	170	175
Glu Ala Tyr Asn Leu Phe Ala Val Asn Gly Ile Leu Phe Asn His Glu	180	185	190
Ser Pro Arg Arg Gly Ala Asn Phe Val Thr Arg Lys Ile Ser Arg Ser	195	200	205
Val Ala Lys Ile Tyr Leu Gly Gln Leu Glu Cys Phe Ser Leu Gly Asn	210	215	220
Leu Asp Ala Lys Arg Asp Trp Gly His Ala Lys Asp Tyr Val Glu Ala	225	230	235
Met Trp Leu Met Leu Gln Asn Asp Glu Pro Glu Asp Phe Val Ile Ala	245	250	255
Thr Gly Glu Val His Ser Val Arg Glu Phe Val Glu Lys Ser Phe Leu	260	265	270
His Ile Gly Lys Thr Ile Val Trp Glu Gly Lys Asn Glu Asn Glu Val	275	280	285
Gly Arg Cys Lys Glu Thr Gly Lys Val His Val Thr Val Asp Leu Lys	290	295	300
Tyr Tyr Arg Pro Thr Glu Val Asp Phe Leu Gln Gly Asp Cys Thr Lys	305	310	315
Ala Lys Gln Lys Leu Asn Trp Lys Pro Arg Val Ala Phe Asp Glu Leu	325	330	335

Val Arg Glu Met Val His Ala Asp Val Glu Leu Met Arg Thr Asn Pro
 340 345 350

Asn Ala

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 372 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Ala His Ala Pro Ala Arg Cys Pro Ser Ala Arg Gly Ser Gly Asp
 1 5 10 15

Gly Glu Met Gly Lys Pro Arg Asn Val Ala Leu Ile Thr Gly Ile Thr
 20 25 30

Gly Gln Asp Gly Ser Tyr Leu Ala Glu Phe Leu Leu Glu Lys Gly Tyr
 35 40 45

Glu Val His Gly Ile Val Arg Arg Ser Ser Ser Phe Asn Thr Gly Arg
 50 55 60

Ile Glu His Leu Tyr Lys Asn Pro Gln Ala His Ile Glu Gly Asn Met
 65 70 75 80

Lys Leu His Tyr Gly Asp Leu Thr Asp Ser Thr Cys Leu Val Lys Ile
 85 90 95

Ile Asn Glu Val Lys Pro Thr Glu Ile Tyr Asn Leu Gly Ala Gln Ser
 100 105 110

His Val Lys Ile Ser Phe Asp Leu Ala Glu Tyr Thr Ala Asp Val Asp
 115 120 125

Gly Val Gly Thr Leu Arg Leu Leu Asp Ala Val Lys Thr Cys Gly Leu
 130 135 140

Ile Asn Ser Val Lys Phe Tyr Gln Ala Ser Thr Ser Glu Leu Tyr Gly
 145 150 155 160

Lys Val Gln Glu Ile Pro Gln Lys Glu Thr Thr Pro Phe Tyr Pro Arg
 165 170 175

Ser Pro Tyr Gly Ala Ala Lys Leu Tyr Ala Tyr Trp Ile Val Val Asn
 180 185 190

Phe Arg Glu Ala Tyr Asn Leu Phe Ala Val Asn Gly Ile Leu Phe Asn
 195 200 205

His Glu Ser Pro Arg Arg Gly Ala Asn Phe Val Thr Arg Lys Ile Ser
 210 215 220

Arg Ser Val Ala Lys Ile Tyr Leu Gly Gln Leu Glu Cys Phe Ser Leu
 225 230 235 240
 Gly Asn Leu Asp Ala Lys Arg Asp Trp Gly His Ala Lys Asp Tyr Val
 245 250 255
 Glu Ala Met Trp Leu Met Leu Gln Asn Asp Glu Pro Glu Asp Phe Val
 260 265 270
 Ile Ala Thr Gly Glu Val His Ser Val Arg Glu Phe Val Glu Lys Ser
 275 280 285
 Phe Leu His Ile Gly Lys Thr Ile Val Trp Glu Gly Lys Asn Glu Asn
 290 295 300
 Glu Val Gly Arg Cys Lys Glu Thr Gly Lys Val His Val Thr Val Asp
 305 310 315 320
 Leu Lys Tyr Tyr Arg Pro Thr Glu Val Asp Phe Leu Gln Gly Asp Cys
 325 330 335
 Thr Lys Ala Lys Gln Lys Leu Asn Trp Lys Pro Arg Val Ala Phe Asp
 340 345 350
 Glu Leu Val Arg Glu Met Val His Ala Asp Val Glu Leu Met Arg Thr
 355 360 365
 Asn Pro Asn Ala
 370

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 bases
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGATGAGCCA GAGGACTTTG TCATAGCTAC

30

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 bases
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CAGAAAGTCC ACTTCAGTCG GTCGGTAGTA

30

19